DOCUMENT RESUME

ED 268 179 TM 860 245

AUTHOR Norcross, John C.; Stevenson, John F.
TITLE Evaluating Clinical Training: Measurement and Utilization Implications from Three National Studies.
PUB DATE Oct 85
NOTE 16p.; Paper presented at the Annual Meeting of the Evaluation Research Society (Toronto, Ontario, Canada, October 17-19, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Clinical Psychology; *Clinical Teaching (Health Professions); *Educational Quality; Evaluation Methods; Higher Education; *Medical Education; National Surveys; *Program Effectiveness; *Program Evaluation; Questionnaires

ABSTRACT

This paper draws on three national studies of professional psychology programs to identify measurement and utilization implications for the evaluation of clinical training. The focus is on the evaluation of the training program rather than the evaluation of students' performance/competence. The directors of 315 programs--74 psychology training clinics, 62 clinical psychology doctoral programs, and 179 predoctoral internships--responded to mailed questionnaires. The questionnaires were designed to: (1) assess the current use of informal and formal training evaluation procedures; (2) gauge the relative impact of these procedures in judging the quality of training; and (3) delineate the major obstacle in conducting this type of evaluation. Aggregate results indicated that evaluation practices heavily favor student-focused, impressionistically-collected, and qualitatively-oriented sources of evidence to judge the training enterprise. Program directors rated written evaluations, formal accreditation reports, trainees' evaluations, and numerical ratings of program effectiveness the most useful measures in gauging the quality of training. Rigor, resistance and resources consistently emerged as the most intractable problems to meaningful evaluation. Findings also indicated considerable convergence in training evaluation designs, measures and obstacles across settings. These marked similarities suggest that site variation is quite limited. Several implications for training evaluation are presented. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED268179

Evaluating Clinical Training: Measurement and Utilization Implications from Three National Studies

John C. Norcross
University of Scranton

John F. Stevenson
University of Rhode Island

Paper presented at the 1985 Evaluation Research Society (ERS) Conference, Toronto, Canada.

TM 860 245

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. C. Norcross

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

Summarizes three national studies of professional psychology programs (total $\underline{N}$ = 315) in order to identify measurement and utilization implications for the evaluation of clinical training. Aggregate results indicated that evaluation practices heavily favor student-focused, impressionistically-collected, and qualitatively-oriented sources of evidence to judge the training enterprise. Program directors rated written evaluations, formal accreditation reports, trainees' evaluations, and numerical ratings of program effectiveness the most useful measures in gauging the quality of training. Rigor, resistance, and resources consistently emerged as the most intractable problems to meaningful evaluation. Findings also suggested considerable convergence in evaluation practices across training programs; site variation seems quite limited. Several implications for training evaluation are presented, including the need for program evaluation, standardized measures, and ongoing internal evaluation designs.

Evaluating Clinical Training:
Measurement and Utilization Implications from Three National Studies

The competence of clinical practitioners and, indeed, the adequacy of clinical training are typically assumed rather than verified (Malouf, Haas, & Forah, 1983; Sechrest & Hoffman, 1982). Local considerations, national priorities, and external demands attest to the need for systematic evaluation of clinical training. Unfortunately, little evaluation of professional skill training programs has been conducted (Bernstein, 1982; Edelstein & Brasted, 1984; Garfield, 1977; Kniskern & Gurman, 1980; Matarazzo, 1978; Miller, 1977; Robertson, 1986). Ford (1979, p. 113), in his review of the research, concluded that "... most counselor/clinician training programs have not evaluated their impac on trainees, nor their trainees' impact on clients, in any systemati and valid fashion."

This paper draws on three national studies of professional psychology programs to identify measurement and utilization implications for the evaluation of clinical training. Our focus is on the evaluation of the training program itself, conceptually distinct (but perhaps pragmatically inseparable) from the evaluation of students' performance/competence.

The directors of 315 programs -- 74 psychology training clinics (Stevenson & Norcross, 1985), 62 clinical psychology doctoral programs (Norcross & Stevenson, 1984), and 179 predoctoral internships (Norcross, Stevenson, & Nash, in press) -- responded to mailed questionnaires. The questionnaires were

designed to: (1) assess the current use of informal and formal training evaluation procedures; (2) gauge the relative impact of these procedures in judging the quality of training; and (3) delineate the major obstacle in conducting this type of evaluation.

Six tables (adapted from Stevenson & Norcross, 1986) summarize the results across the three surveys. The first 5 tables group sources of evidence for training quality on the basis of three dimensions: focus of evaluation (aggregating performance indicators for individuals vs. assessing the program directly); type of evaluation data (qualitative vs. quantitative); and method of data collection (impressionistic vs. systematic). Without exception, results of our three studies indicated that training evaluation practices heavily favor student-focused, impressionistically-collected, qualitatively-oriented sources of evidence to judge the training enterprise.

The top measures of rated impact were supervisors' written evaluations, the APA accreditation report, trainees' evaluations of supervisors, supervisors' evaluations of trainees, and numerical ratings of program effectiveness. Notable by their low impact ratings were formal quantitative designs and empirical across-program comparisons.

Three "R's" consistently emerged as the most intractable problems to meaningful evaluation in professional psychology program (see Table 6). These were Rigor (e.g., technological limitations, inadequate measures), Resistance (e.g., noncompliance, apathy, overt resistance), and Resources (e.g., time restraints,

inadequate funding). Similar obstacles have been encountered by others involved in the evaluation of clinical training (e.g., APA, 1982; Grant & Anderson, 1977).

Several significant correlates of evaluation policy impact were identified across programs. Internal financial support, ongoin routine evaluation, and the number of measures employed were positively related to training directors' rated impact of training evaluation on program policy.

Inspection of the tables indicates considerable convergence in training evaluation designs, measures, and obstacles across settings. These marked similarities suggest that site variation is quite limited.

Major implications:

(1) Evaluation of clinical training is in a pre-paradigmatic stage characterized by immense variability in measures, dissensus about proper focus (e.g., student vs program, formative vs summative), and paucity of appropriate technology.

(2) A second and more positive implication concerns the need for greater involvement of program evaluators (see also Levy, 1983). Training directors, by their own admission, require assistance in constructing standardized measures, creating sensitive designs, and providing evaluation resources.

(3) Further, given the limited variation in evaluation practices attributable to program settings, improvements should equally benefit all health-related programs in evaluating

their clinical training.

(4) Finally, for those interested in initiating or improving their training evaluation, these aggregate results provide several useful directions.

(a) Encourage evaluations of trainees, supervisors, and the program, preferably in a written, structured, and in some cases, anonymous format.

(b) Invite colleagues to visit your program and to offer candid feedback.

(c) Include some numerical ratings or standardized measures to compliment the pervasive use of impressionistically-collected, qualitatively-oriented methods.

(d) Adopt an ongoing, internal evaluation design. The self-evaluating organizations is highly-prized; clinical trainers find little value in rigorous comparative studies at this time.

References

American Psychological Association Task Force on Education, Training, and Service in Psychology. (1982). Summary report. Washington, DC: Author.

Bernstein, G.S. (1982). Training behavior change agents: A conceptual review. Behavior Therapy, 13, 1-23.

Edelstein, B.A., & Brasted, W.S. (1983). Clinical training. In M. Hersen, A.E. Kazdin, & A.S. Bellack (Eds.), The clinical psychology handbook. New York: Pergamon.

Ford, J.D. (1979). Research on training counselors and clinicians. Review of Educational Research, 49, 87-130.

Garfield, S.L. (1977). Research on the training of professional psychotherapists. In A.S. Gurman & A.M. Razin (Eds.), Effective psychotherapy: A handbook of research. New York: Pergamon.

Grant, D.L., & Anderson, S.B. (1977). Issues in the evaluation of training. Professional Psychology, 8, 659-677.

Kniskern, D.P., & Gurman, A.S. (1980). Research on training in marriage and family therapy: Status, issues, and directions. In M. Andolfi & I. Zwerling (Eds.), Dimensions of family therapy. New York: Guilford Press.

Levy, L.H. (1983). Evaluation of students in clinical psychology programs: A program evaluation perspective. Professional Psychology: Research and Practice, 15, 497-508.

Malouf, J.L., Haas, L.J., & Forah, J. (1983). Issues in the preparation of interns: Views of trainers and trainees. Professional Psychology: Research and Practice, 14, 624-631.

Matarazzo, R.G. (1978). Research on the teaching and learning of psychotherapeutic skills. In S.L. Garfield & A.E. Bergin (Eds.), Handbook of psychotherapy and behavior change (2nd ed., pp. 941-966). New York: Wiley.

Miller, P.M. (1977). Evaluation of trainee performance in psychology internship programs. The Clinical Psychologist, 30, 2-5.

Norcross, J.C., & Stevenson, J.F. (1984). How shall we judge ourselves? Training evaluation in clinical psychology programs. Professional Psychology: Research and Practice, 15, 497-508.

Norcross, J.C., Stevenson, J.F., & Nash, J.M. (in press). Evaluation of internship training: Practices, problems, and prospects. Professional Psychology: Research and Practice.

Sechrest, L., & Hoffman, P.E. (1982). The philosophical underpinnings of peer reveiw. Professional Psychology, 13, 14-18.

Stevenson, J.F., & Norcross, J.C. (1985). Evaluation activity in psychology training clinics: National survey findings. Professional Psychology: Research and Practice, 16, 29-41.

Stevenson, J.F., & Norcross, J.C. (1986). Current status of training evaluation in clinical psychology. In B. Edelstein & E. Berler (Eds.), Evaluation and accountability in clinical training. New York: Plenum.

Robertson, M. (1986). Training eclectic psychotherapists. In J.C. Norcross (Ed.), Handbook of eclectic psychotherapy. New York: Brunner/Mazel.

Table 1

Use and Importance of Informal Qualitative Measures

% Use				Importance Rating[a]	
Intern Programs	Grad Programs	Training Clinics	Measure	Intern Programs	Grad Programs
99%	95%	--	Your own impressions of students' growth	4.3	3.9
99%	95%	91%	Supervisors' verbal evaluations of students' clinical work	4.4	4.1
100%	90%	--	Your impressions of students' satisfaction with the program	4.2	3.6
82%	84%	--	Subjective contrast of your program with others with which you are familiar	3.2	2.9
91%	86%	--	Reputation of your training program	3.8	3.5
87%	87%	--	Reputation of your former students among their professional peers	3.8	3.9

[a]Importance ratings are based on a 5-point Likert-type scale (1 = not important to 5 = very important)

Table 2

Use and Importance of Formal Aggregated Measures

% Use				Importance Rating[a]	
Intern Programs	Grad Programs	Training Clinics	Measure	Intern Programs	Grad Programs
93%	92%	82%	Written evaluations of students by clinical supervisors	4.6	4.5
61%	60%	--	Numerical ratings of students by clinical supervisors	3.9	4.1
34%	47%	--	Measures of clinical problem-solving skills (e.g., simulated cases)	3.5	3.3
45%	26%	--	Students' self-ratings of ability	3.5	3.0
8%	13%	--	Simulated licensure examination	2.7	3.3
8%	--	--	Simulated ABPP model exam	3.3	--
14%	58%	16%	Tests of factual knowledge (outside of specific course/ seminar requirements)	3.0	4.2
55%	58%	--	Systematic evaluation of videotaped therapy sessions	3.7	3.9
76%	69%	--	Systematic evaluation of case presentations	4.1	3.8
85%	81%	--	Systematic evaluation of written clinical reports	4.3	4.0

[a]Importance ratings are based on a 5-point Likert-Type scale (1 = not important to 5 = very important).

Table 3

Use and Importance of Formal Comparative Designs

% Use			Importance Rating[a]	
Intern Programs	Grad Programs	Measure	Intern Programs	Grad Programs
29%	26%	Longitudinal (e.g., pre-post training) within group comparison	3.5	3.2
26%	39%	Cross-sectional comparison of new students with those who have already received training	3.1	2.8
6%	19%	Formal comparison to students in other programs	2.3	2.8

[a]Importance ratings are based on a 5-point Likert-style scale (1) = not important to 5 = very important).

Table 4

Use and Importance of Post-Training Measures

% Use			Importance Rating[a]	
Intern Programs	Grad Programs	Measure	Intern Programs	Grad Programs
67%	95%	Quality of internships that students obtain	3.5	4.0
20%	24%	Systematically obtained feedback from employers	3.3	3.1
48%	52%	Rate of successful certification/ licensure of former students	3.6	3.2
36%	--	Proportion of former students holding formal leadership positions (e.g., officer, committee chair) in professional organizations	3.2	--
15%	90%	Proportion of former students achieving ABPP Diplomate status	3.2	3.2
24%	--	Publication rate of former students	3.1	--

[a]Importance ratings are based on a 5-point Likert-type scale(1 = not important to 5 = very important).

Table 5

Use and Importance of Program-Focused Measures

% Use			Importance Rating[a]	
Intern Programs	Grad Programs	Measure	Intern Programs	Grad Programs
80%	89%	Quantitative evaluation of didactic training by students (e.g., seminar/ course evaluations)	4.2	4.1
81%	74%	Quantitative evaluation of supervisors by supervisees	4.4	4.2
83%	73%	Systematic site visits to evaluate training	4.3	4.1
73%	45%	Quantitative ratings by students of program effectiveness	4.3	3.7
87%	76%	Type and number of seminar/course offerings	4.0	3.6
76%	53%	Student quantitative satisfaction with clinical training	4.2	3.7
51%	26%	Quantitative ratings by former students of program effectiveness	4.0	3.4
39%	23%	Quantitative ratings by faculty of program effectiveness	3.8	3.0
100%	92%	APA Accreditation Report	4.5	4.3
99%	--	Program planning activity with intern input	4.3	--

[a]Importance ratings are based on a 5-point Likert-type scale (1 = not important to 5 = very important).

Table 6

Problems in Evaluating Training

Problem	Mean Rating		Frequency
	Intern Programs	Graduate Programs	Training Clinics
Time restraints	2.7	2.8	28
Inadequate funding	3.4	3.4	15
Lack of personnael	3.1	3.1	9
Transience of students and problems with foloow-up	3.3	3.2	6
Heterogeneity of students	4.0	4.0	5
General apathy and disinterest	4.0	3.6	20
Student resistance, noncompliance	4.2	4.1	12
Faculty resistance, noncompliance	3.8	3.8	--
Philosophical and theoretical disparities among faculty	4.2	4.1	12
Lack of belief in the value of research evidence for this purpose	3.9	3.9	3
Ethical considerations	4.3	4.2	2
Inadequate evaluation methods and measures	3.0	3.0	19

[a]Based on a 5-point Likert-type scale (1 = very serious obstacle to 5 = not an obstacle)